

# SELF TIMED BIT AND READ/WRITE PULSE STRETCHERS

## BACKGROUND OF THE INVENTION

5

### *Field of the Invention*

The present invention is related to random access memories (RAMs) and more particularly to static RAM (SRAM) access timing.

10

### *Background Description*

15

Integrated circuits (ICs) are commonly made in the well-known complementary insulated gate field effect transistor (FET) technology known as CMOS. A typical CMOS circuit includes paired complementary devices, i.e., an n-type FET (NFET) paired with a corresponding p-type FET (PFET), usually gated by the same signal. Since the pair of devices have operating characteristics that are, essentially, opposite each other, when one device (e.g., the NFET) is on and conducting (ideally modeled as a closed switch), the other device (the PFET) is off, not conducting (ideally modeled as an open switch) and, vice versa. For example, a CMOS inverter is a series connected PFET and NFET pair that are connected between a power supply voltage ( $V_{dd}$ ) and ground (GND).

20

A typical static random access memory (SRAM) cell is a pair of cross coupled inverters storing a single data bit. A pair of pass gates (FETs) selectively connect the complementary outputs of the cross coupled inverter to a corresponding complementary pair of bit lines. A word line connected to the gates of the pass gate FETs selects connecting the cell to the corresponding complementary pair of bit lines. Normally, an  $N$  row by  $M$  column SRAM array is organized as  $N$  word lines by  $M$  column lines. Each column line includes one or more ( $K$ ) bit line pairs. Accessing  $K$  bits (for a read or a write) from array entails driving one of the  $N$  word lines, turning on the pass gates for all

YOR920030497US1

1

Express Mail Label No.

ER588459778US

Date of Deposit  
December 15, 2003

M by K cells on that word line. With the pass gates on for that selected word line, the cross coupled cell inverters are coupled to the corresponding bit line pairs, partially selecting the M by K cells on that word line. Selection of one of the M columns selects the K cells on that word line, the Kbits actually being accessed. The remaining (M – 1) 5 by K bits remain partially selected during the access. During a read, each partially selected cell couples its contents to the corresponding bit line pair such that each of the bit line pairs rises/droops, usually, only to develop sufficient signal (e.g., 50mV) for a sense amplifier. The selected K bit line pairs are coupled to a sense amplifier, which senses the contents of the selected cells from the signal on the coupled K bit line pairs. 10 Then, after sensing data for the selected Kbits, the word line returns low again, deselecting/isolating the M by K cells on that word line.

During a write, however, the K selected bit line pairs are driven to opposite extreme voltages ( $V_{dd}$  and GND) or write voltages with the bit line voltages for the 15 remaining partially selected cells being substantially the same as for a read access. With the write voltages on the selected bit line pairs and the word line high, the write voltages on the bit line pairs begin to pass through the selected cell pass gates, i.e. to the cell cross coupled inverters. Any selected cell that is being written with what it already stores, remains unchanged. Any selected cell that is in the opposite state of what is being 20 written must be switched, which takes a minimum time depending upon the cell design and cell technology known as the cell write time. For an ideally balanced cell, it is sufficient to force the cross coupled latches just beyond the voltage mid points (i.e., to  $V_{dd}/2 + \delta/2$  and  $V_{dd}/2 - \delta/2$ ) or beyond cross over before dropping the word line and allowing the cross coupled latches to switch the rest of the way. So, once cell voltages 25 cross over, the word line may be dropped to isolate the M by K cells from the bit line pairs and to capture the new data in the cells. Once the word line is low, the bit line pairs may be released, e.g., both of each pair driven or restored high and decoupled from the write driver.

If insufficient signal develops (i.e.,  $<\delta$ ) in the cell, however, , the data write may fail and, the cell may remain unswitched or become meta-stable. Either result is unsatisfactory and unreliable because cell contents are indeterminate. So, the write may fail, for example, if the word line drops too soon or, the bit line pair voltage change too soon, e.g., from the write driver terminating prematurely. To avoid this and insure that each write is successful, both the word line must be held high for the minimum write time and, the selected bit line pairs must be held at the write voltages at least until after the word line is returned low.

10 For a synchronous SRAM design, typically, word selection is a multiple of a timing period, e.g., a half cycle, chosen to meet array timing constraints. So, for a write, while the word line is selected for that multiple, i.e., at least as long as the minimum write time, a second longer timing unit (e.g., 2 timing periods or a full cycle) are required for bit and write control signals to insure that the bit line pair voltages remain stable until 15 after the word line is unselected. This extends the write access time. Unfortunately, once sufficient additional time is added for restoring the bit lines and write driver, access cycles are considerably longer than the word line select, perhaps as much as three or four times as long. This impairs SRAM performance and performance for anything accessing the SRAM.

20 Thus, there is a need to reduce RAM access time.

## SUMMARY OF THE INVENTION

25 It is a purpose of the invention to improve RAM data reliability;

It is another purpose of the invention to minimize RAM write access time;

It is yet another purpose of the invention to insure data is written reliably to selected cells in a minimized write access time.

The present invention relates to bit and write decode/drivers, a random access memory (RAM) including the decode/drivers and an IC with a static RAM (SRAM) including the decode/drivers. The decode/drivers are clocked by a local clock and each produce access pulses wider than corresponding clock pulses. The bit decode/driver produces bit select pulses that are wider than a word select pulse and the write decode/driver produces write pulses that are wider than the bit select pulses for stable self timed RAM write accesses.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

The foregoing and other objects, aspects and advantages will be better understood from the following detailed description of a preferred embodiment of the invention with reference to the drawings, in which:

Figure 1A shows an example of a block diagram of a preferred embodiment memory with a high performance self timed bit decode and write pulse stretcher;

Figure 1B shows a timing diagram for the memory example of Figure 1A;

Figure 2A shows an example of a bit decode pulse stretcher;

Figure 2B shows a timing diagram for the bit decode pulse stretcher of Figure 2A;

Figure 2C shows an example of a column select driver for a complementary bit line pair;

Figures 3A – B show an example of a preferred embodiment write pulse stretcher and a corresponding timing diagram.

## DESCRIPTION OF PREFERRED EMBODIMENTS

5

Referring now to the drawings, and more particularly, Figure 1A shows an example of a block diagram of a memory 100, e.g., a random access memory (RAM) macro or chip, with a high performance self timed bit decode and write pulse stretcher, according to a preferred embodiment of the present invention. Figure 1B shows a timing diagram for the memory example of Figure 1A. In this example, the memory array 102 includes cells of well known six transistor (6T) latches or storage cells or 8T 2 port RAM cells (not shown) organized in N rows of word lines by M columns of K bit lines. More particularly, the storage array may be a typical CMOS SRAM or 2 port SRAM in what is known as silicon on insulator (SOI) technology, although application of the present invention is advantageous to almost any technology and any SRAM.

Cell selection is by coincidence a column selected by preferred bit decode and select circuit 104 with a row selected by word decoder 106. Selected cells are coupled to suitable state of the art sense amplifiers 108 for reading data stored in cells during a read. Data from the sense amplifiers 108 are passed to suitable state of the art data input/output (I/O) transceivers 110. Clock logic 112 provides local timing. A write pulse stretcher 114 selectively enables self timed array writes, synchronized by the clock logic 112. Data in for a write selectively passes from I/O transceivers to cells in the array 102 as selected by the bit decode and select circuit 104 and enabled by write pulse stretcher circuit 114. Glue logic (not shown) provides local control logic.

As can be seen from the timing diagram example of Figure 1B, showing the relationship of various signals to a local clock 115 from clock logic 112 providing local timing synchronization edges. Word decoder 106 provides N word line signals 116. Bit

decode and select circuit 104 provides M bit select signals 117. A Read/Write (RW) input 118 to write pulse stretcher 114 initiates a write pulse 119 from the write pulse stretcher 114. It should be noted that the timing edges are not to scale and representative of the positional timing relationships only. With each clock cycle (115), one of the N word lines (116) may be pulsed high, selecting a corresponding one of the N word lines, with the remaining N-1 word lines held low. Also, one of the M bit select signals 117 may be selected in each access, with the remaining M-1 bit select signals held low. The bit select pulses in 117 are longer than the word line pulses in 116, which insures that at the end of each access, the cells on the selected word line are isolated from corresponding bit lines before the bit line states change. Assertion of a WRITE 118 initiates a write select pulse 119, that is longer than the bit select pulse (117) and insures that selected bit lines are disconnected from the bit write driver (e.g., in Data I/O 110) before the write pulse ends. Thus, a preferred embodiment memory provides a self timed write, stretched such that it is only marginally longer than the word line pulse width.

15

Figures 2A – C show an example of a cross section of a preferred bit decode and select circuit 104 of Figure 1 and timing for the cross section. Figure 2A shows an example of a bit decode pulse stretcher 120 that includes an address decode 122, a pulse stretcher 124 and a driver 126. In this example the address decode 122 includes one of eight decode logic 128, e.g., a dynamic NOR decode receiving a 3 bit partial address (b0, b1, b2) signal at the gates of parallel connected NFETs 128-1, 128-2, 128-3, which are connected between a common source node 134 and a decode node 136. It should be noted that, although the address decode 122 of this example is a one of eight dynamic NOR decode, this is for example only and not intended as a limitation. Any suitable decode logic may be used, including but not limited to, self resetting logic or delayed clock logic.

20  
25  
A decode precharge PFET 138 gated by pulse stretcher 124 precharges decode node 136 high. Pulse stretcher 124 includes a 2 input NAND gate 140 and delay 142. In

this example, delay 142 is a group of (4) series connected inverters 144, 146, 148, 150. The clock 115 is the input to delay 142 and one input to the 2 input NAND gate 140. The output 152 of delay 142 is the second input to 2 input NAND gate 140. The output 154 of 2 input NAND gate 140 is the output of pulse stretcher 124 and drives decode  
5 precharge PFET 138, decode enable NFET 156 and decode enable precharge PFET 158. Decode enable NFET 156 is connected between common source node 134 and a supply return, e.g., ground. Decode enable PFET 158 is connected between the decode output 160 and a supply voltage ( $V_{dd}$ ). Decode node 136 is connected to the gate of NFET 162 and to the drain of PFET 164. NFET 162 is connected between decode output 160 and common source node 134. The decode output 160 also is connected to the gate of pseudo  
10 latch PFET 164, which is connected between  $V_{dd}$  and decode node 136 and holds decode node 136 high when it is left floating high, i.e., the particular bit line is not selected. The decode output 160 is the input to driver 126, which includes a driver NFET 166, a pseudo latch PFET 168 and, in this example, a pair of driver PFETs 170A and 170B driving output 172, i.e., a column select. Thus, driver NFET 166 is connected between the output  
15 172 and ground; and, PFETs 170A and 170B are connected between the output 172 and  $V_{dd}$ . Pseudo latch PFET 168 is connected between  $V_{dd}$  and decode output 160 and is gated by output 172, e.g., 118 in Figure 1B.

20 As can be seen form the timing diagram of Figure 2B for one of the eight (in this example) decoders 120, at steady state between accesses with the clock input 115 high, the delay output 152 is high and pulse stretcher output 154 is low. The low on pulse stretcher output 154 holds PFETs 138 and 158 on, clamping decode node 136 and decode output 160 both high. NFET 156 is off, floating common source node 134. With decode  
25 output 160 high, driver output 172 is low and pseudo latch PFET 166 is off. With output 172 low, pseudo latch PFET 166 is on, pulling decode output 160 high. A decode occurs on the fall of the clock input 115, when the output of a single selected decoder 120 is driven high.

5

So, in this example, each of the three address signals, b0, b1 and b2, is a true or complement of one of three address bits. Except for the selected address decoder 122 at least one of these three bit address signals, b0, b1 and b2, rises or is high for all but one address decoder 122, i.e., the address decoder 122 corresponding to the selected column address. So, when the clock input 115 falls, the pulse stretcher output 154 rises, turning off precharge PFET 138 and turning on decode enable NFET 156 which pulls common source node 134 to ground. For the seven (in this example) unselected bit address decoders 122, the decode node 136 is pulled low, holding NFET 162 off. With NFET 162 off, decode output 160 remains high and bit decode output 172 remains low.

10

15

20

25

For the selected address decoder 122, however, the decode node 136 remains high. So, NFET 162 turns on, pulling decode output 160 low, which turns on pseudo latch PFET 164 to clamp the decode node 136 high. In response to the low on the address decoder output 160, the driver 126 drives bit decode output 172 high, which is the complement of the address decoder output 160. With bit decode output 172 high, pseudo latch PFET 168 turns off. When the clock low period ends and the clock 115 rises, pulse stretcher 154 remains high until the clock edge passes through the delay 142. When the rising edge of the clock exits the delay 142, both inputs to NAND gate 140 are high to drive the pulse stretcher output 154 low. The low on pulse stretcher output 154 turns off decode enable NFET 156 and turns on decode precharge PFET 138 and decode enable precharge PFET 158. Decode precharge PFET 138 pulls the decode node 136 high on the seven unselected decoders 122 with the eighth remaining high. Decode enable precharge PFET 158 pulls the selected decoder output high 160 and, in response, the driver 126 drives output 172 low; the unselected seven outputs remain low. Thus, the pulse out of the selected decoder output 172 is approximately the same width as the pulse stretcher output 154 of NAND gate 140 and, longer than both the word line pulse and the clock low period, stretched by the length of the delay 142.

Figure 2C shows a column select driver 180 for a complementary bit line pair

182, 184, connected to a number (N) of cells (not shown), each connected to one of N word lines in an array 102. The column select driver 180 includes a pair of series connected inverters 186, 188. The first inverter 186 receives a decoded column select signal 172 from a preferred embodiment bit decode pulse stretcher, e.g. 120. The second inverter 188 drives bit line pull up devices, PFETs 190, 192, and an equalization device, PFET 194. The output of the first inverter 186 is an input to a 2 input NOR gate 196 and drives a pair of bit line select pass gates, PFETs 198, 200, which are read pass gates, passing a complementary signal on the selected bit line pair 182, 184 to a sense amplifier (108 in Figure 1) during a read on complementary data line pair 202, 204, respectively. 5 A write control signal 119 is a second input to the 2 input NOR gate 196. A pair of write devices, NFETs 208, 210, are driven by the output 212 of 2 input NOR gate 196, selectively coupling complementary input data on data write pair 214, 216 to bit line pair 182, 184, respectively. 10

15 In a typical access, an array word line (not shown) is driven high selecting a row of cells and, a selected column signal 172 pulses high at the input to the corresponding first inverter 186 to select one column. The output of the first inverter 186 falls and the output of the second inverter 188 rises. The high turns off bit line pull up devices 190, 192 and equalization device 194, floating the bit line pair 182, 184, allowing a signal to 20 develop. The low on bit line select pass gates 198, 200 couples the bit line pair 182, 184 to the data line pair 202, 204. During a read, the write input 119 to NOR gate 196 remains high. So, the write devices 208, 210 remain off because the output of NOR gate 196 is low. During a write, the write input 119 pulses low. So, the write devices 208, 210 turn on when the output first inverter falls, driving the output of NOR gate 196 high. 25 With the write devices 208, 210 on, data passes from data write pair 214, 216 to the bit line pair 182, 184.

Figure 3A show an example of a write pulse stretcher 114, which includes a pulse stretcher 222, a READ/WRITE decode 224 and a driver 226 and Figure 3B is a

corresponding timing diagram. As with pulse stretcher 124 of Figure 2A, this pulse stretcher 222 also includes a 2 input NAND gate 228 and a delay 230. In this example, delay 230 is a group (8) of series connected inverters 232, 234, 236, 238, 240, 242, 244, 246. This delay 230 operates substantially the same as bit decode delay 142 in Figure 2A except that delay 230 stretches the write pulse by approximately twice the amount as bit decode delay 142. The same clock 115 is an input to the 2 input NAND gate 228 and delay 230. The output 248 of the delay 230 is the second input to NAND gate 228. The output of NAND gate 228 is the output 250 of the pulse stretcher 222 and drives READ/WRITE decode 224. It should be noted that any suitable delay may be selected; provided, that the pulse width is such that the trailing (falling) edge exits before the end of the clock up period and does not encroach on the next following clock, which could result in double pulsing.

Continuing this example, the READ/WRITE decode 224 is a dynamic NOR with a PFET/NFET complementary pair 252, 254 series connected between  $V_{dd}$  and a write enable node 256 and a pair of parallel connected NFETs 258, 260 between write enable node 256 and ground. It should be noted that, although both address decode logic 130 in Figure 2A and READ/WRITE decode 224 are shown herein as NOR gates, any suitable decode logic may be substituted. NFET 258 is gated by a write select signal 118 and NFET 260 is gated by a test write signal, e.g., for loading the array during test. The drains of the complementary pair 252, 254 are the output 262 of the READ/WRITE decode 224 and the input to the driver 226. The driver 226 includes a pseudo latch PFET 264 and a pair of series inverters 266, 268. The pseudo latch PFET 264 is connected between  $V_{dd}$  and the READ/WRITE decode output 262 and is gated by the output 270 of the first inverter 266. The output of the second driver inverter 268 is the write pulse stretcher output 119.

At steady state between accesses, when the clock 115 is high, the delay output is high and pulse stretcher output 250 is low. The low on pulse stretcher output 250 holds

NFET 254 off and PFET 252 on to pull decode output 262 high. With decode output 262 high, the output 270 of inverter 266 is low, driver output 119 is high and pseudo latch PFET 264 is on. As noted hereinabove, delay 230 operates substantially identically as described for bit decode pulse stretcher 124. So, when the clock 115 is low, pulse  
5 stretcher output 250 is high; when the clock 115 rises, pulse stretcher output 250 falls, but only after the clock traverses the delay 230 ; and, when the clock 115 falls again, the pulse stretcher output 250 rises with no additional delay. With both write select signals low to parallel NFETs 258, 260, READ/WRITE decode output 262 and driver output 119 remain high; inverter 266 holds pseudo latch PFET 264 on, clamping READ/WRITE  
10 decode output 262 high. Thus, regardless of the clock state, unless either of the write select signals is high, write pulse stretcher output 119 remains high. However, when either of the write select signals is high, READ/WRITE decode 224 acts as an inverting driver, passing the low clock pulse through the pulse stretcher 222, which stretches the pulse as described above for bit decode 120.

15

Advantageously, bit decode pulses are wider than word line pulses; and write pulses are longer than bit decode pulse. Therefore, provided the word line select pulse is long enough for a write, data is reliably written with each write and without appreciably extending the write access beyond a read access. Thus, the present invention improves  
20 SRAM performance and reliability, providing maximum available read and write times without compromising array cell stability, especially for half selected cells. In particular, the trailing edges of the bit select and write pulses overlap the word select pulse, which may be as little as 40% of the minimum cycle time. Further, the present invention has application to any suitable RAM, e.g., a 2 port RAM, wherein a write takes an  
25 appreciably longer time than a read.

While the invention has been described in terms of preferred embodiments, those skilled in the art will recognize that the invention can be practiced with modification within the spirit and scope of the appended claims.